

TITLE OF INVENTION

Enhanced Engineered Chromosome Formation from Alpha Satellite with Artificially
Increased Density of CENP-B Boxes

5

BACKGROUND OF THE INVENTION1. Field of the Invention

The invention relates to the field of artificial chromosomes and gene expression. It is demonstrated that using alpha satellite DNA containing an increased number of CENP-B boxes enhances the efficiency of *de novo* artificial chromosome formation.

10

2. Background of the Related Art

15

Alpha satellite DNA is the major species of repetitive element found at the centromeres of all normal primate chromosomes. It is organized in a hierarchical structure based on a ~171 bp monomeric unit that is multimerized in a tandem manner into a higher-order repeat, which is further multimerized over hundreds to thousands of kilobases at the centromeres of all normal human chromosomes (reviewed in 1, 2, 3, 4). Centromeric alpha satellite acts to organize the recruitment of key centromeric proteins (CENPs) to form a trilaminar protein/DNA complex, the kinetochore. The kinetochore mediates the interactions between the chromosome and the spindle apparatus that are responsible for coordinated chromosome movements during cell division (5). While functional kinetochores have been observed at chromosomal locations not containing any alpha satellite (so called "neo-centromeres"; reviewed in (6)), only cloned alpha satellite DNA has thus far been shown to form centromeres *de novo* when introduced into the cell

20

nucleus by transfection or microinjection in synthetic microchromosome (SMC) assays (7, 8, 9).

The ability to create SMCs *de novo* was pioneered through the development of techniques to synthesize extended length alpha satellite arrays *in vitro*, including megabase size synthetic arrays (10), starting with a single cloned copy of a higher-order repeat (11). These SMCs are useful as vectors in gene transfer (7,12); for example, SMCs containing the *HPRT* genomic locus have been shown to complement HPRT-deficient cell lines (Rudd et al., manuscript in preparation, 13, 14), and the present inventors have observed sustained expression of the β -globin gene from SMCs carrying the entire 150 kb β -globin genomic region (Basu et al., in preparation). In addition, SMC and artificial chromosome vectors provide a methodological platform for the identification and functional analysis of elements in alpha satellite that are critical for centromere function (Rudd et al. (Nov 2003), *Mol. Cell. Bio.* 23(21):7689-7697; also see 15, 16, 17, 10).

SUMMARY OF THE INVENTION

The presence of binding sites for the centromere protein CENP-B (the 'CENP-B box') has been correlated with the ability of alpha satellite DNA to form centromeres *de novo* in synthetic microchromosome (SMC) assays. However, the effect of the density of CENP-B boxes on the frequency of SMC formation has not previously been explored. The present disclosure reports a systematic analysis of the role of the CENP-B box in human alpha satellite DNA, using the formation of SMCs as an assay for the establishment of centromere function. Synthetic alpha satellite arrays were created based on the 16-

monomer repeat length typical of natural chromosome 17-derived D17Z1 arrays. In these synthetic arrays, the consensus CENP-B box elements were either completely absent (0/16 monomers) or were increased in density (16/16 monomers) compared to D17Z1 alpha satellite (5/16 monomers). The test results demonstrated that the presence of CENP-B
5 box element is required for efficient de novo centromere formation and that increasing the density of CENP-B box elements in the alpha satellite DNA results in enhancement of the efficiency of de novo centromere formation. These findings have implications for the design of strategies to construct novel SMC vectors for functional genomics and potential therapeutic applications.

10 Accordingly, a first embodiment of the present invention relates to an engineered higher order repeat DNA comprising one or more CENP-B boxes, wherein said one or more CENP-B boxes are distributed on the engineered higher order repeat DNA in an order other than that of CENP-B boxes on a naturally-occurring higher order repeat DNA.

A second embodiment of the invention relates to an engineered alphoid DNA
15 comprising one or more CENP-B boxes, wherein said one or more CENP-B boxes are distributed on the alphoid DNA in an order other than that of CENP-B boxes on a naturally occurring alphoid DNA.

Hence, certain embodiments of the invention relate to engineered higher order repeat DNA and/or alphoid DNA enriched in CENP-B box sequences.

20 Other embodiments of the invention relate to engineered chromosomes and chromosome vectors containing the alphoid DNA or the HOR DNA that is enriched in CENP-B box quantity and/or order. Yet another embodiment relates to an engineered

chromosome formed by introduction of the above-mentioned engineered chromosome into an appropriate cell.

5 In a preferred embodiment of the invention, when the engineered chromosome vector enriched in CENP-B boxes is introduced into an appropriate cell it forms an engineered chromosome at an efficiency rate greater than an engineered chromosome vector containing a higher order repeat DNA with a naturally-occurring frequency or distribution order of CENP-B boxes.

10 A most preferred embodiment of the invention relates to an engineered chromosome vector enriched in its number of CENP-B boxes, wherein said engineered chromosome vector forms an engineered chromosome upon introduction into an appropriate cell at an efficiency rate of greater than about 1-5%, about 5-15%, about 10-20%, or about 15-25% compared to a corresponding engineered chromosome vector which is not enriched in its number of CENP-B boxes.

15 Yet another preferred embodiment of the invention relates to an engineered chromosome enriched in its number of CENP-B boxes, wherein said engineered chromosome is mitotically stable inside an appropriate cell. A most preferred embodiment of the invention relates to a mitotically stable engineered chromosome with a mitotic segregation pattern that is substantially 1:1.

20 Another most preferred embodiment of the invention is an engineered chromosome vector comprising a transposon.

Another embodiment of the invention relates to a method of increasing efficiency of formation of an engineered chromosome containing alphoid DNA comprising adding

one or more CENP-B boxes to the alphoid DNA used to form said engineered chromosome.

A further embodiment of the invention relates to a method of making an alphoid DNA array comprising constructing two or more engineered monomers of defined DNA sequences; wherein at least one monomer is enriched in CENP-B box sequences; and assembling said engineered monomers to form said alphoid DNA array. Accordingly, an embodiment of the invention relates to an engineered alphoid DNA array made by this process.

Yet another embodiment of the invention relates to a method of making an engineered higher order repeat DNA comprising constructing two or more engineered monomers of defined DNA sequences; wherein at least one monomer is enriched in CENP-B box sequences; and directionally assembling said engineered monomers to form said higher order repeat DNA.

Accordingly, a further embodiment of the invention relates to a higher order repeat DNA made by the above method.

A further preferred embodiment of the invention relates to a method of engineering a desired higher order repeat DNA comprising engineering each monomer unit of said desired higher order repeat DNA as one or more oligonucleotide(s); wherein at least one monomer is enriched in CENP-B box sequences; and directionally ligating pairs of adjacent monomer units to form repeating monomeric units to form the desired higher order repeat DNA.

Accordingly, a further preferred embodiment of the invention relates to a higher order repeat DNA made by the above method.

A most preferred embodiment of the invention relates to an engineered chromosome vector, wherein said vector when introduced in an appropriate cell forms an engineered chromosome at an efficiency rate higher than an engineered chromosome
5 vector containing higher order repeat DNA with fewer CENP-B boxes.

Additional advantages, objects, and features of the invention will be set forth in part in the description which follows and in part will become apparent to those having ordinary skill in the art upon examination of the following or may be learned from practice of the
10 invention. The objects and advantages of the invention may be realized and attained as particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in detail with reference to the following drawings:

15 **Figure 1(A)** depicts an outline of an iterative scheme for synthesis of mutant versions of chromosome 17 alpha satellite arrays. Each of the 16 individual monomers comprising a single higher-order repeat (HOR) was synthesized as 2-3 oligonucleotide pairs (60-100 bp each), which were directly ligated together and gel purified. Adjacent repeat units were subsequently ligated to form dimers as shown and PCR-modified to introduce
20 SapI recognition sites at both ends as appropriate. Digestion with SapI allows seamless ligation of adjacent dimers to create tetramers without introduction of extraneous non-alpha satellite sequences. Two additional rounds of serial ligation resulted in formation of a

complete synthetic higher-order repeat unit, which was subcloned into the BAC vector pBeloBAC (Shizuya et al., 1992), creating pBAC17 α 1(all CENP-B box+/all CENP-B box-).

Figure 1(B) depicts an outline of a scheme for directional multimerization of engineered higher-order repeats. A synthetic alpha satellite array consisting of 32 tandemly multimerized copies of the higher-order repeat was created as follows: pBAC17 α 1 was digested with BamHI and SpeI and the alpha satellite containing fragment (fragment 'A') isolated and gel purified. The same construct was separately digested with BglII and SpeI, and the larger fragment (fragment 'B') isolated and gel purified. Ligation of fragment 'A' to fragment 'B' is directional, resulting in head-to-tail multimerization of adjacent higher-order repeats. The resulting pBAC17 α 2 construct was then isolated following transformation of the ligation reaction into *E.coli*. This process was repeated iteratively to create the final pBAC17 α 32 arrays.

Figure 1(C) depicts pulsed Field Gel Electrophoresis (PFGE) analysis of intermediates in the construction of 17 α 32 HOR/BeloBAC constructs. Each intermediate was digested with NotI, which excises the entire subcloned alpha satellite array from the pBeloBAC vector backbone. Lanes are labeled according to higher-order repeat copy number. The insert in lane 4 is 2.7 kb and therefore too small to be resolved by PFGE in the example shown.

Figure 2 depicts mobility shift analysis of synthetic CENP-B box enriched and CENP-B box null monomers. Ligated tetramers of CENP-B box-enriched and CENP-B box-null monomers were electrophoresed through an agarose gel following incubation with

purified recombinant CENP-B protein. Lanes 1, 2, and 3 represent enriched tetramers, while lanes 4, 5, and 6 contain null species. Tetramer DNAs (100ng) were pre-incubated with varying quantities of CENP-B protein for 25 minutes at room temperature and subsequently loaded into a 2% agarose gel. Lanes 2 and 5 (20µg protein) as well as lanes 3 and 6 (40µg protein) contain protein/DNA mixtures. Comparison of lanes 2 and 3 to lanes 5 and 6 reveals a marked difference in mobility shift in the CENP-B box-enriched subunits, while only a modest shift is seen with CENP-B box-null DNA. This slight mobility shift is likely due to salt effects as similar results are observed with a buffer-only control (data not shown).

Figure 3 depicts cytogenetic detection of SMCs from synthetic chromosome 17-derived alpha satellite arrays. Arrows designate SMCs. Immunostaining with an anti-CENP-C antibody (green) identifies functional centromeres. FISH with the synthetic alpha satellite as probe (red) hybridizes with the synthetic microchromosome as well as to the centromeres of the endogenous chromosome 17s. DAPI stained DNA is shown in blue.

Figure 3(A) depicts generating HT1080 clone by transfection with pBAC17α32(All CENP-B box+), showing the presence of two SMCs.

Figure 3(B) depicts generation of HT1080 clone by transfection with pBAC17α32(natural). A single SMC is visible.

Figure 3(C) depicts generation of HT1080 clone by transfection with pBAC17α32(CENP-B Box null). Two putative SMCs are present in this clone, but none were detected in all other clones obtained with the CENP-B box null construct.

Figure 4 depicts a transposon vector for rapid retrofitting of genomic BACs into unimolecular BAC-SMC vectors. The 86 kb 17 α 32HOR alphoid array was subcloned as a BamH1/Bgl2 fragment into the BamH1 site of the transposon vector. This implies that digestion of a genomic BAC with BamH1 will indicate the approximate size of the alphoid array inserted therein. Tel=telomere; ME=transposase recognition site; 17 α 32HOR=32 copies of the 2.7 kb Higher Order Repeat derived from the centromere of chromosome 17; P_{gk}-puro=puromycin resistance cassette; Neo/Kan=dual neomycin/kanamycin resistance marker.

Figure 5 depicts the molecular analysis of unimolecular BAC-SMC vectors.

Figure 5 (A) depicts a schematic of the BAC-SMC vector used to generate these microchromosomes. An 86 kb synthetic chromosome 17 derived alpha satellite array is marked with XXXX. The solid thin black line marks the 10 kb BAC vector backbone. Digestion of a SMC generated from this BAC with I_{ce}u-1 is predicted to generate a single, discrete band of approximately 100 kb, as seen in Figure B, lane 1, and Figure C, lanes 1 and 3. Digestion with Asc1 and Mlu1 generates an 86 kb alpha satellite containing insert and a 10 kb vector dropout, as seen in Figure B, lane 2 and Figure C, lanes 2 and 4.

Figure 5 (B) depicts a Southern blot analysis of HT1080 clone containing SMC. Lane 1: I-Ceu1 digest of the genomic DNA plugs. Lane 2: Asc-1/Mlu1 digest of the genomic DNA plugs. Digests were separated by PFGE, transferred and hybridized with a BAC vector backbone specific probe.

Figure 5 (C) depicts PFGE analysis of BACs rescued by Hirt extraction of HT1080 clones identified by southern analysis. Lanes 1-2: are I-ceu1 and Asc1/Mlu1 digests of a typical clone. Lanes 3-4: are I-ceu1 and Asc1/Mlu1 digests of another example.

Figure 6 also depicts molecular analysis of unimolecular BAC-SMC vectors.

5 **Figure 6 (A)** depicts a genomic plug southern, HT1080 control (lane 1) and clone G6B -1 (lane 2) cut with I-Ceu1. The band of 200 kb is the linear form of the original G6 B+ head to head construct.

Figure 6 (B) depicts a Hirt gel, colonies generated from the Hirt prep and transformation into bacteria. Digested colony from clone G6 B-1 (lanes 3 and 4), I-Ceu1 and Not 1, respectively. Control G6 B+ head to head DNA cut with I-Ceu1 and Not 1 (lanes 1 and 2).

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

It is well recognized that the formation of a functional centromere is at the heart of making synthetic chromosomes. A characteristic of primate, including human, centromeres is that it is composed of a major class of repetitive DNA known as alpha satellite DNA. This DNA, also referred to as alphoid DNA, is comprised of a monomeric repeating unit of about 171 bp. These monomeric units are organized into different tandem arrays that constitute clearly definable higher-order repeating (HOR) structures or alphoid subfamilies. Numerous (at least about 33) different alphoid subfamilies or HOR structures have been identified to date. Some of these HOR structures are specific for a single naturally-occurring chromosome, while others are common to a group of naturally-

15

20

occurring chromosomes. Moreover, some chromosomes appear to have only a single HOR structure within their centromeres, whereas others may be comprised of several different HOR structures. More detailed information is available and known to skilled artisans regarding the genomic distribution of alpha satellite DNA on all human chromosomes. For example, such information, including a derived evolutionary consensus sequence for alpha satellite monomer consensus sequence is additionally provided by K.H. Choo *et al.* in *Nuc. Acid Research*, Vol. 19, No. 6, pp.1179-1182 (1991), which is herein incorporated by reference in its entirety. It should be noted that alphoid DNA is highly polymorphic, and such polymorphic sequences, as well as mutants (especially silent mutants), are useful in the practice of the present invention.

All terms pertaining to recombinant DNA technology are used in their art-recognized manner and would be evident to one of ordinary skill in the art.

Appropriate cell: refers to a cell (e.g., mammalian, primate, human) that allows formation therein of an engineered chromosome.

Y alpha satellite and Y α : are used interchangeably and refer to alpha satellite DNA derived from the Y chromosome.

17 alpha satellite and 17 α : are used interchangeably and refer to alpha satellite DNA derived from chromosome 17.

Alphoid (DNA), alphoid (DNA) monomer, monomer repeats: Alphoid DNA is the only repetitive satellite DNA sequence found in the centromeric region of primate, e.g. human, chromosomes. In humans, the size of the array on each chromosome varies between approximately 78 kb and 5 Mb (*see*, Yang, J.W., *et al.* (2000) Human mini-

chromosomes with minimal centromeres. *Hum. Mol. Genet.* **9**: 1891-1902). Alphoid DNA consists of 171 base pair monomer repeats organized into larger higher-order repeat (HOR) units. There are at least 12 distinct monomer types, classified into five suprachromosomal families according to the organization of the monomer units (*see*, Lee, C., *et al.* (1997) Human Centromeric DNAs *Hum. Genet.* **100**:291-304). For example, 17 α belongs to suprachromosomal family 3 and consists of type W monomers repeated as a pentamer (W1-5) forming a characteristic HOR. Y α , on the other hand, is classified in suprachromosomal family 4 and has a monomeric organization without a distinctive HOR and exhibits only the type M alphoid monomers of this family. The other family 4 chromosomes (for example, 21), however, belong to other alphoid suprachromosomal families as well, and contain monomers of the D and/or R types in addition to M; *see* Lee (1997).

Centromere: the region of the chromosome that is constricted and is the site of attachment of the spindle during meiosis or mitosis. It is necessary for the stability and proper segregation of chromosomes during meiosis and mitosis and is therefore an essential component of artificial chromosomes. Centromeric DNA comprises a DNA that directs or supports kinetochore formation and thereby enables proper chromosome segregation. Centromeric DNA at active, functional, centromeres is associated with CENP-E during mitosis, as demonstrated by immunofluorescence or immunoelectron microscopy. By "associated" is meant that the centromeric DNA and CENP-B co-localize by FISH and immunofluorescence.

CENP-B Box: The CENP-B box is the stretch of DNA, present on alphoid DNA monomers from all human chromosomes except Y. It is minimally responsible for mediating binding of the constitutive centromere protein CENP-B to human alpha satellite DNA. At present, the biochemically-defined 17-bp degenerate sequence motif
5 “5’PyTTCGTTGGAAPuCGGGA3’” is a structure determined in the art to be capable of providing this binding function (20, 21). For example, 5’aTTCGttggAaaCGGGa3’ is a typical CENP-B box sequence, wherein the bases indicated by capital letters are the most important for the binding of the CENP-B protein, whereas the bases indicated by lower case letters may be substituted with other bases.

10 **Directionally** as in **directionally ligating**: refers to the order of the fragments that are ligated together in a sequential order, following the sequence of the DNA unit that is being constructed. For example, in constructing a fragment with the following sequence “ATTTTITAGCGCCCGGTTTATTTACCCCCCCC,” the smaller fragments that are first constructed span the full length of the larger fragment. For example, 4 smaller fragments
15 may be constructed with the following sequences: Fragment 1 = ATTTTITA; Fragment 2 = GCGCCCGG; Fragment 3 = TTTATTTA; and Fragment 4 = CCCCCCCC. By “directionally ligating” the smaller fragments, therefore, it is meant that small fragment 1 is ligated to small fragment 2 and the small fragment 3 is ligated to small fragment 4, all in the same sequential orientation 5’ to 3’ or 3’ to 5’, to maintain the sequence of the larger
20 fragment that is to be constructed. It would NOT be “directionally ligating” if fragment 1 were to be ligated to fragment 3 or 4 and/or if the 5’ to 3’ direction of the sequence of any one small fragment was disrupted (as in ligating the small fragment 1 in its 5’-3’ direction to

the small fragment 2 in its 3'-5' direction, resulting in a larger fragment with the sequence ATTTTTTA + GGCCCGCG, instead of the directional ligation sequence of ATTTTTTA + GCGCCCGG).

Engineered Chromosome (EC): refers to any form of episomal vectors whether
5 obtained by the so-called "bottom-up" or "top-down" methodologies. The bottom-up approach aims to assemble a new chromosome *de novo* from its constituent DNA elements, and the product is commonly referred to as an artificial or synthetic chromosome or microchromosome. The "top-down" approach starts with an existing human chromosome, which then becomes experimentally reduced in size to a minichromosome.
10 For convenience, the products generated by both these strategies are referred to collectively as engineered chromosomes (ECs). The minimum components that a successful EC needs to have are: (1) sequence motifs or structural elements (such as hairpin loops) that signal DNA replication, necessary for the self-propagation of the chromosome; (2) a centromere, which is essential for the accurate segregation of the replicated sister chromatids to
15 daughter cells; and (3) telomere sequences at both ends of a linear chromosome (not necessary for non-linear, e.g. circular chromosomes) to provide structural stability to the chromosome ends. Most genomic DNA pieces larger than 20-30 kb carry some origins of replication, an EC of a size around or greater than a mega-base (Mb) should, therefore, typically contain these motifs.

20 **EC vector:** denotes a non-naturally occurring chromosome vector regardless of how it is made.

Endogenous DNA: denotes DNA naturally contained within a given cell as opposed to any DNA that might have been introduced into the cell from the outside, such as a vector DNA.

Engineered: As opposed to naturally-occurring, “engineered” denotes man-made or -designed. For example, an engineered HOR denotes a DNA molecule with the repetitive sequence structure of a higher order repeat DNA unit; wherein unlike the naturally occurring HOR, the engineered HOR is made using any suitable laboratory technique (chemical synthesis, isolation from nature, full or part amplification of the DNA, site-directed mutagenesis, recombination and breakage of naturally-occurring or synthetic DNA, ligation of two or more DNA fragments, or any combination of methodologies known to the skilled artisans for making the molecule). HOR DNA may be considered engineered whether because it has been fully or partially synthesized, expressed, constructed, or assembled *de novo* or because it has been obtained by altering the natural centromeric alpha-satellite DNA of a naturally-occurring chromosome or of an engineered chromosome derived from a naturally occurring chromosome.

Enriched: denotes an increase in a quantity. For example, “enriched in CENP-B box sequences” denotes an increase in the number of CENP-B box unit sequences compared to a corresponding DNA unit containing fewer number of CENP-B boxes (e.g., a naturally-occurring HOR or an engineered chromosome with fewer CENP-B boxes, etc).

Essential chromosome functions: include mitotic stability without experimental selective pressure, substantially 1:1 segregation, autonomous replication, i.e., centromere, telomere (for linear chromosomes), and origin of replication functions.

Exogenous DNA: denotes DNA introduced into a cell from outside. Another copy of the same DNA may already exist in the cell, which would be called the endogenous copy of that DNA.

Heterologous: a DNA sequence not found in the naturally-occurring genome of the cell in which the artificial mammalian chromosome is introduced. Additionally, if the sequence is found in the genome of the cell, any additional copies that might be discovered in the cell upon transfection are considered "heterologous" because they are not found in that form in the naturally-occurring genome.

Higher Order Repeat (HOR) unit: HOR, as described above, refers to a repeating unit of DNA that is itself composed of smaller (monomeric) repeating units also referred to as alphoid (DNA) monomers (*see Alphoid DNA*, above). Monomers are organized into chromosome-specific higher order repeating units, which are also tandemly repetitive. The number of constituent monomers in a given HOR varies, from as little as two (for example, in human chromosome 1) to greater than 30 (human chromosome Y). Constituent monomers exhibit varying degrees of homology to one another, from approximately 60% to virtual sequence identity. However, HORs retain a high degree of homology throughout most of a given alphoid (DNA) array.

Isolated: refers to DNA that has been removed from a cell.

Isoschizomer refers to a restriction enzyme that recognizes the same nucleotide sequence as another restriction enzyme and cleaves that same sequence. Therefore, a "Non-isoschizomeric site" refers to a restriction enzyme site that can be cut by one of two restriction enzymes, but not by both.

Mammalian chromosome: means a DNA molecule or genetic unit that functions as a chromosome in a mammalian cell.

Naked DNA: means DNA that is unassociated with any of the biological (chromosomal or cellular) components with which it is normally associated in a naturally-
5 occurring chromosome, for example histones, non-histone chromosomal proteins, RNA, transcription factors, topoisomerases, scaffold proteins, centromere-binding proteins, and telomere-binding proteins. Such DNA can be isolated from cells and purified from the non-DNA chromosomal components. Alternatively, this DNA can be synthesized *in vitro*.

Naturally-occurring: denotes events or objects that occur in nature and are not
10 experimentally-induced or made.

Non-naturally occurring distribution of CENP-B boxes: denotes a structural arrangement within a HOR unit that differs from a naturally-occurring HOR in that the number (including absence thereof) and/or the position of the CENP-B boxes has been altered from the natural arrangement. In the present invention, both the distribution of the
15 CENP-B boxes as well as the number of CENP-B boxes may be altered to form a desired DNA construct. For example, a construct may contain a CENP-B box in every HOR or one in every other HOR, or none in the first 5 HOR, and so on and so forth. Such constructs are useful *per se* (as for example, increasing efficiency of SMC formation) or useful in a variety of ways in the elucidation of the role of various permutations of the
20 molecular structure of HORs in centromere formation and function. Both increases and decreases in the efficiency of artificial chromosome formation may be desired in order to achieve a particular effect, for example, control gene expression levels.

Origin of replication: a site or region of initiation of DNA synthesis.

Purified DNA: refers to isolated DNA that has been substantially completely separated from non-DNA components of a cell or to DNA that has been synthesized *in vitro* and separated substantially completely from the materials used for synthesis that would interfere with the construction of the chromosome from the DNA. A purified DNA can also be a DNA sequence isolated from the DNA sequences with which it is naturally associated.

Replicon: a segment of a genome in which DNA is replicated and by definition contains an origin of replication.

Seamless restriction enzyme: any restriction enzyme that would allow ligation of two DNA fragments of a higher repeat order DNA (such as the pairs of adjacent dimers shown in Figure 1A) to form a larger fragment (such as the tetramers shown in Figure 1A) without introduction of extraneous non-alpha satellite sequences. Examples of "Seamless" enzymes include the class of restriction enzymes known as Type IIS. Type IIS enzymes like FokI and AlwI cleave outside of their recognition sequence to one side. These enzymes are often of intermediate size, typically 400-650 amino acids in length, and they recognize sequences that are continuous and asymmetric. They comprise two distinct domains, one for DNA binding, the other for DNA cleavage. They are thought to bind to DNA as monomers for the most part, but to cleave DNA cooperatively, through dimerization of the cleavage domains of adjacent enzyme molecules. For this reason, some Type IIS enzymes are much more active on DNA molecules that contain multiple recognition sites. Restriction enzymes that cleave sites that occur naturally in the HOR

may also be used to ligate fragments of synthetic alphaoid DNA that have been modified as described herein, such as the enrichment or elimination of CENP-B boxes in a monomer or assembled HOR.

Synthetic centromeric alpha-satellite DNA: denotes a DNA molecule with the repetitive sequence structure of a centromeric alpha-satellite DNA, wherein the synthetic centromeric alpha-satellite DNA is made using any laboratory technique (such as chemical, recombinant DNA methodology) suitable for obtaining a centromeric alpha-satellite DNA. Centromeric alpha-satellite DNA may be considered synthetic whether because it has been fully or partially synthesized, expressed, constructed, or assembled *de novo* or because it has been obtained by altering the natural centromeric alpha-satellite DNA of a naturally-occurring chromosome.

Synthetic or artificial chromosome: are used interchangeably. A "synthetic" or "artificial" chromosome is a construct that has essential chromosome functions but which is not naturally-occurring. It has been created by introducing exogenous DNA into a cell. Since the chromosome is composed entirely of exogenous DNA, it is referred to as synthetic or artificial. A **synthetic microchromosome** more precisely points out that the size of the artificial chromosome is smaller than a natural chromosome (generally, that is because it does not carry as many exons and introns as a natural chromosome). However, a synthetic microchromosome is an artificial or synthetic chromosome, and an artificial or synthetic chromosome may be made as large or larger than the naturally-occurring chromosomes if desirable.

Engineered Chromosome (EC) vector: sometimes used interchangeably with Engineered Chromosome (EC) denotes a polynucleotide molecule, made by man (i.e., non-naturally occurring), having the minimum structural requirements for forming a functional chromosome. As long as this polynucleotide has not been yet interacted with other factors
5 (such as the required proteins - whether outside or inside a cell) to form a functional SMC, it is preferably referred to as an EC vector. However, once it is functional and behaving as a chromosome, it is more appropriately referred to as an EC. Nevertheless, it should be noted that an EC is useful as a vector for delivering, propagating, and/or expressing other desired DNA; hence, continuing to preserve its use as a vector even after it has become a
10 functional engineered chromosome. For example, an EC can act as a vector when a desired DNA sequence is transposed onto it.

Mitotic stability: as used with regard to a chromosome, such as an EC or SMC, denotes the structural integrity and segregation pattern of such chromosome inside an appropriate cell after at least about 30 generations of growth with a low or non-existent
15 recombination frequency. The preferred ranges of recombination frequency are less than about 0.5 per generation for an EC vector that contains at least 32 copies of natural or CENP-B box enriched HOR. Similarly, the mitotic segregation pattern of the EC in human cells is substantially 1:1, meaning that during each cell division of a cell that contains a single copy of the EC, there is better than about a 95% probability that each
20 daughter cell will receive a single copy of the EC. Preferably, there is a better than 99% probability that each daughter cell will receive a single copy of the EC.

Transfecting or transforming: as used interchangeably herein denotes the introduction of nucleic acids into a cell. The nucleic acid thus introduced is not naturally in the cell in the sequence introduced, the physical configuration, or the copy number.

Telomere: denotes the end of a chromosome comprising simple repeat DNA that is synthesized by a ribonucleoprotein enzyme called telomerase. The function is to allow the ends of a linear DNA molecule to be replicated and structurally stabilized.

The inventors have observed variations in the efficiency of *de novo* centromere formation between alpha satellite templates derived from different human chromosomes (18, 8, 16), and have proposed that a causal link exists between the presence of sequence elements called CENP-B boxes and *de novo* centromere seeding efficiency (15, 19).

While there was clear evidence implicating the presence of CENP-B boxes in *de novo* centromere formation (15), it was not clear to what extent the density of CENP-B boxes might influence the efficiency of SMC formation. Thus, in order to address the functional significance of the CENP-B box in human alpha satellite and in SMC formation, the inventors developed methodologies to directly vary the density and distribution of CENP-B boxes in the D17Z1, chromosome 17-derived HOR, which in its natural configuration contains a CENP-B box in 5 of its 16 constituent monomers. Hence, entirely synthetic D17Z1 HOR derivatives were constructed, in which each of the 16 tandem monomeric repeats contains either a consensus CENP-B box or a related sequence element derived from Y chromosome alpha satellite, which does not bind CENP-B (22, 23). It was observed and is herein reported that the efficiency of formation of SMCs is directly proportional to the density of CENP-B boxes in the SMC vector, thus demonstrating a

requirement for CENP-B boxes in centromeric chromatin assembly. As the methods presented here are generally applicable, these data have implications for the design and further development of SMCs for potential applications in protein production as well as human gene therapy.

5 Since the original report of *de novo* centromere and SMC formation (10), a number of groups have described related approaches to further develop and optimize artificial chromosome systems (reviewed by 7, 26, 12). The creation of SMCs has now been established as a tractable approach to systematically identify and dissect elements that are critical for chromosome function (15, 16, and Rudd et al. (Nov 2003), *Mol. Cell. Bio.*
10 23(21):7689-7697). The present disclosure describes, *inter alia*, the further refinement of the SMC system as a methodological platform to undertake a functional analysis of the role of the density of CENP-B box elements in human alpha satellite DNA.

CENP-B is a constitutively present DNA-binding protein found in the underlying centric heterochromatin of all human chromosomes except the Y chromosome. The
15 corresponding DNA sequence element that defines the cognate binding site, the CENP-B box, has been identified as PyTTCGTTGGAAPuCGGGA (20, 22) and is found distributed within some, but not all, of the monomer units of alpha satellite DNA from most human centromeres (25, 16, 27). However, the role of CENP-B if any, in specifying centromeric identity globally remains unsettled (28). Y chromosome centromeres do not
20 associate with CENP-B (23), and African Green Monkey centromeres lack CENP-B boxes even though the CENP-B protein itself is present (29). Furthermore, *Cenp-B* knockout

mice show only modest phenotypic effects and appear to have fully functional centromeres as evidenced by the lack of chromosome mis-segregation phenotypes (30, 31, 32).

Notwithstanding this mechanistic uncertainty, studies of *de novo* centromere formation with cloned alpha satellite arrays support a direct correlation between the presence of CENP-B boxes and the competence of a construct for *de novo* centromere formation. For example, comparison of cloned alpha satellite arrays from chromosomes Y, X, 17 and 21 show that 17- and 21-derived arrays form *de novo* centromeres much more efficiently than X- and Y-derived arrays (Rudd et al. (Nov 2003), *Mol. Cell. Bio.* 23(21):7689-7697; and 8, 18). In addition, alpha satellite from a CENP-B box rich region of the chromosome 21 centromere (21-I) forms *de novo* centromeres in an SMC system, while alpha satellite from a neighboring CENP-B box depleted region (21-II) is inefficient (19). Further, the *de novo* centromere nucleation ability of the 21-I-derived alpha satellite array can be disrupted by mutation of its constituent CENP-B boxes (15), an outcome that parallels the presently presented observations on mutation of CENP-B boxes in D17Z1-derived alpha satellite. Finally, it has also been established that CENP-B boxes outside the context of alpha satellite DNA are not competent to nucleate *de novo* centromere assembly (15), establishing that sequence features other than CENP-B boxes are also required for centromere function. Taken together, the presently disclosed data and the earlier observations unambiguously establish the presence of CENP-B and its cognate binding element as a requirement for efficient *de novo* centromere formation in SMC or artificial chromosome assays.

Despite the clear role of the CENP-B box in assembly of SMCs, the role of CENP-B in its endogenous chromosomal context remains open to debate. At least three CENP-B-like proteins have been identified in fission yeast, and double mutants exhibit severe chromosome segregation defects (33). Such functional redundancy may explain the lack of a major phenotype in mouse knockouts of *Cenp-B* (29, 30, 31) and why *Cenp-B* appears dispensable for function of the Y chromosome in both mice and men, as well as for function of neocentromeres and certain dicentric chromosomes (34, 35). In addition, it remains to be established whether the position of CENP-B boxes within an array of monomers or even within a single monomer is also of importance, as might be expected if CENP-B participates in nucleosome positioning (36, 37).

In addition to the effect of manipulating CENP-B boxes demonstrated here and by Ohzeki et al. (15), it is apparent that other sequences within alpha satellite may influence the efficiency of SMC formation, as even arrays with a similar number of CENP-B boxes can differ quite substantially in their ability to seed SMCs (Rudd et al. (Nov 2003), *Mol. Cell. Bio.* 23(21):7689-7697; and 25). This possibility may now be investigated systematically using synthetic alpha satellite arrays where the distribution of CENP-B boxes and/or other sequences in each monomer has been manipulated, using the approach outlined here. Determination of the ideal density and distribution of such sequences in alpha satellite will maximize the efficiency with which SMC vectors carrying therapeutic genes might eventually be assembled in human cells (14, 7, 12).

The methodology described in the examples of the present disclosure for making the synthetic HOR and alpha satellite array as well as the synthetic artificial mammalian

chromosome is preferred. However, any modifications of the disclosed methodology or other methodologies known to the skilled artisans may be used as well. The manipulation of efficiency of chromosome formation is achieved by altering the density and distribution of the CENP-B box and it is not critical how this objective is achieved. For example,
5 USPN 5,695,967 (Van Bokkelen *et al.*), which is incorporated in its entirety herein by reference, provides detailed description of a method for making repeating tandem arrays of DNA which is useful in making the synthetic HORs and the synthetic centromeric alpha satellite DNA of the present invention. USPN 6,348,353 B1 (Harrington *et al.*), which is incorporated in its entirety herein by reference, sets forth a preferred method of making
10 artificial mammalian chromosomes that are useful for making the claimed invention.

A general preferred approach for building up the array is to start with a construct such as the pBeloBAC17alpha X HOR CENP-B box saturated/null. X is the number of copies of the HOR in a given iteration. X may equal 1, 2, 4, 8, 16, 32 copies of the approximately 2.6 kilobase HOR, etc. Taking the embodiment where $X = 1$, as shown in
15 Figure 1B, digestion of the starting construct with BamH1 and Spe1 creates an insert fragment, referred to as "A," consisting of the HOR plus a small amount of vector sequence. Digestion of the starting construct with Bgl2 and Spe1 creates the corresponding vector fragment or "B," consisting of the starting vector minus the small amount of sequence between the Bgl2 and Spe1 sites. A is now cloned into B to give the
20 pBeloBAC17alpha2HOR, shown on the right, in Figure 1B. Reiteration of this process builds up the array to pBeloBAC17alpha32HOR and so forth.

The CENP-B box sequences used in the present invention may be isolated from alpha-satellite DNA of a given chromosome or it may be fully or partly synthesized chemically and the partial DNA sequences maybe ligated together by any means known in the art in order to form a CENP-B box DNA unit.

5 A preferred embodiment of the invention is directed to increasing the frequency of formation of an engineered chromosome, e.g., SMC, by increasing the number of CENP-B boxes present on the centromeric alphoid DNA array. The frequency rate may be increased by any percentage or fraction thereof, for example, by greater than about 5-10%, 10-15%, 15-20%, 20-25% frequency rate of EC formation of a corresponding EC vector
10 differing only in that it contains fewer CENP-B boxes. Hence, the preferred embodiment of the present invention enables making engineered chromosome vectors with improved frequency rate of formation of engineered chromosomes, e.g., a SMC.

The preferred engineered chromosome of the invention is mitotically stable, meaning that it is capable of being propagated in an appropriate host cell for at least about
15 30 generations of growth with a low or non-existent recombination frequency. The preferred ranges of recombination frequency are less than about 0.5 per generation for an EC vector that contains at least 32 copies of natural or CENP-B box enriched HOR. Similarly, the mitotic segregation pattern of the EC in human cells is substantially 1:1, meaning that during each cell division of a cell that contains a single copy of the EC, there
20 is better than about a 95% probability that each daughter cell will receive a single copy of the EC. Preferably, there is a better than 99% probability that each daughter cell will receive a single copy of the EC.

The present invention may be preferably practiced by making a transposon vector that contains a synthetic alpha satellite array (either naturally occurring or enriched for CENP-B boxes). Figures 4 depicts an example of such transposon vector which was specifically designed for rapid retrofitting of genomic BACs into unimolecular BAC-SMC
5 vectors. Transposon systems and their use in vector construction are known in the art (*see* for example, Goryshin, I.Y. and Reznikoff, W.S. (1998) *J. Biol. Chem.* 273, 7367 and USPN 5,965,443, herein incorporated by reference, as well as Davies, D.R. *et al.* (2000) *Science* 289 (5476), 77).

Optionally, such transposon vector includes additional elements such as one or
10 more selectable markers, and/or telomeric DNA, as described herein. Such a vector may also be transposed into another plasmid that contains any desired fragment of DNA, for example including human genomic DNA that contains a gene (or multiple genes) of interest. Plasmids that contain the desired gene(s) of interest and the transposon may then be screened and structurally analyzed in order to identify a vector clone that possesses the
15 desired structural configuration (e.g. insertion of the transposon vector into the appropriate region of the target plasmid). In this way, the transposon based approach may be used to rapidly retrofit any BAC vector that contains a DNA construct or fragment of interest, including cloned fragments of human DNA.

Optionally, the transposon vector may also be engineered to include elements that
20 facilitate packaging of newly constructed vectors into viral capsids, such as HSV-1 particles, using a viral amplicon system, such as those described in the literature. Preferably, such

vectors should be of appropriate size so as to be efficiently accommodated into the viral capsid upon vector packaging, as described in the literature.

The present invention may also be practiced by constructing SMC vectors that are packaged into viral capsids using techniques that are known to those skilled in the art (*see* for example E. Antonio Chiocca et al, "Viral delivery Systems for Infectious Transfer of Large Genomic DNA Inserts" Pub. No.: U.S. 2002/0110543 A1 (Pub Date August 15, 2002; and Howard J. Federoff et al, "Helper Virus-Free Amplicon Particles and Uses Thereof" Pub. No.: U.S. 2003/0027322 A1 (Pub Date Feb 6, 2003). Such SMC vectors have a variety of uses such as *in vitro* protein expression or gene therapy.

10 Examples

Previous studies have established that vectors containing multiple copies of certain alpha satellite higher-order repeat units can seed formation of *de novo* centromeres in human HT1080 cells (8, 10, 15-18; Rudd et al. (Nov 2003), *Mol. Cell. Bio.* **23**(21):7689-7697). However, the overall frequency of generation of SMCs has been reported to be quite variable and often quite low (Rudd et al., in press; 25, 15, 8, 18), depending at least in part on the chromosomal origin of the alpha satellite array and on the presence or absence of CENP-B boxes. Therefore, the inventors developed, and herein describe, a general approach to maximize the efficiency of SMC formation and to evaluate the sequence-dependency of *de novo* centromere seeding.

20 Materials & Methods

The following materials and methods were used by the inventors which provide specific teachings as well as general guidelines for making and using the claimed invention.

All the materials & methods as well as the experiments described in the Examples provide sufficient guidance to persons of skill in the art to carry out the invention and are in no way intended to limit the scope of the claims.

5 *Synthesis of modified 2.7 kb chromosome 17-derived higher-order repeats*

The sequence of the 2.7 kb D17Z1 higher-order repeat (11) was modified such that each of the 16 monomer units contained the consensus CENP-B box element 5': TTT CGT TGG AAA CGG GA: 3' (22) or the related Y alpha satellite-derived element AGA TGG TGG AAA AGG AA, which lacks CENP-B-binding activity ('CENP-B box null').

10 Each of the 16 modified monomer units was then synthesized by ligation of two to three pairs of overlapping oligonucleotides (Operon Technologies, CA). Adjacent pairs of mutated monomer units were then ligated together to form dimers. In addition, the EcoRI sites of monomers 1 and 16 were altered to create a BamHI site at the 5' end of monomer 1 and a BglII site at the 3' end of monomer 16. Each gel-purified dimer was then PCR
15 amplified with a BsaI or SapI restriction site, such that upon digestion each dimer would produce a defined overhang exactly complementary to an overhang in the adjacent dimer. The resultant tetramers (containing no extraneous sequence) were then T/A subcloned into pGem-Teasy (Promega) and sequence verified. Adjacent tetrameric subunits were then ligated together using SapI (or NotI and SapI for monomers 1 and 16) to generate the
20 appropriate overhang. The resultant octamers were further gel purified and ligated together to produce the completed synthetic 16-mer, representing a single D17Z1 higher-order repeat unit, with NotI overhangs. This higher-order repeat was then subcloned as a

NotI fragment into the BAC cloning vector pBeloBAC11 (24). The overall strategy is outlined in Figure 1A.

5 *Directional multimerization of the synthetic higher-order repeats*

The 2.7 kb CENP-B box enriched or CENP-B box null D17Z1 higher-order repeat was multimerized directionally as follows. The cloned synthetic higher-order repeat (in pBeloBAC11) was digested with BamHI and SpeI, and this band (fragment 'A') was gel purified by standard procedures (Qiagen). A second fragment ('B') was generated by
10 digesting the same cloned repeat with BglII and SpeI. The appropriate fragment 'B' was subsequently gel purified and ligated to the BamHI/SpeI digested fragment 'A'. This ligation reaction was transformed into *E.coli* (GibcoBRL), and recombinant clones identified by NotI digestion of the resultant clones and pulsed field gel electrophoresis (Fig. 1B). This process was repeated iteratively to create clones containing 4, 8, 16 and 32 copies
15 of the CENP-B box enriched/CENP-B box null chromosome 17 based higher-order repeat in pBeloBAC (Fig. 1C). Finally, for use as a selectable marker in mammalian cells, a cDNA cassette conferring resistance to puromycin was introduced into 17 α 32(CENP-B box enriched/null) unit/pBeloBAC by transposition of the puroR cassette into the pBeloBAC vector backbone (Epicentre).

20 An ~86 kb synthetically assembled alpha satellite array, derived from directional multimerization of the *naturally* occurring 2.7 kb D17Z1 repeat unit (p17H8, see 8, 10, 11), was subcloned as a BamHI/BglII fragment into the BamHI site of pBeloBAC11. This

construct, 17 α 32(natural)/pBeloBAC, was further modified by transposition with a puromycin resistance selectable marker (Epicentre). The structural integrity of all modified higher-order repeats and of the original higher-order repeat array was confirmed by sequencing, restriction digestion and FISH hybridizations using the array as probe.

5

Mobility shift analysis

The effect of mutations described above on CENP-B binding to the synthetic HOR was evaluated by a gel mobility shift assay. Cloned tetramer units assembled from CENP-B box-enriched and CENP-B box-null monomers were digested with NotI and inserts were gel purified. Subsequent to incubation with purified recombinant CENP-B protein (Diarect, Germany) for 25 minutes at room temperature in CENP-B binding buffer (20), protein/DNA complexes were electrophoresed through a 2% agarose gel in 0.5xTBE buffer. Following electrophoresis, SybrGold (Molecular Probes) stain was used to visualize DNA bands.

15

Cell transfection

Human HT1080 cells (gift of Dr. Brenda Grimes, Case Western Reserve University) were transfected using the Eugene 6 (Roche) reagent according to the manufacturer's instructions, and stable clones identified on the basis of resistance to puromycin (Kayla) at 3 μ g/ml. Clones appeared after 7-10 days and were subsequently expanded to generate clonal lines for further analysis.

Cytogenetic analysis and validation of SMCs

Clonal populations of cells containing potential SMCs were analyzed, generally as described (8, 16, 10). Briefly, cells were arrested at metaphase using colchicine (Gibco) at 40 ug/ml for 45 minutes at 37 degrees Celsius, then treated with hypotonic solution (0.075 M KCl, 12 minutes, 37 degrees Celsius) and applied to slides using the Shandon Cytospin 3. Slides were subsequently fixed in 2% formaldehyde solution and immunoreacted with rabbit anti-CENP-C antibody (10) at a concentration of 1/2000 in PBS and detected with goat anti-rabbit IgG (H + L) (Molecular Probes). DNA probes were labeled by nick translation using the Vysis system according to the manufacturer's instructions. Immunoreacted slides were fixed (3:1, methanol:acetic acid), subjected to denaturation (70% formamide, 72 degrees Celsius, 8 minutes), and hybridized to denatured probes as described (8).

Putative artificial chromosomes were scored if they showed a positive hybridization signal with a FISH probe derived from the synthetic array as well as positive CENP-C immunoreactivity. Mitotic stability was evaluated by growth in the absence of drug selection for up to six weeks.

Construction of Engineered, D17Z1-based higher-order repeats

The SMC system provides a platform to systematically evaluate the functional significance of sequence elements within human alpha satellite DNA. The inventors developed methodologies to construct modified synthetic D17Z1 units that are either enriched or depleted in the density of CENP-B box DNA binding elements. The higher-

order repeat unit of D17Z1 alpha satellite consists of 16 monomer units (11). In order to generate engineered higher-order repeats, each of the 16 monomer units was synthesized by the serial stepwise assembly of oligonucleotide pairs, each between 60 and 100 bp in length, as shown in Figure 1A. Adjacent monomer units could then be gel-purified and
5 ligated to form dimers. Each dimer was PCR-amplified to introduce a restriction site such as SspI (which cuts outside its recognition sequence and can generate custom-made overhangs that can be ligated seamlessly), thereby generating tetramers without the addition of any extraneous sequence. This process of PCR and ligation assembly was serially repeated until the complete 16-mer repeat unit was constructed. The resulting synthetic
10 higher-order repeat was then subcloned and directionally concatamerized to 32 copies (Figure 1B, C), using methods previously developed in the inventors' lab (10).

CENP-B boxes are required for efficient centromere formation *de novo*

The inventors used the approach described above to create a modified variant of
15 D17Z1 alpha satellite in which all the consensus CENP-B boxes or elements resembling the consensus in each of the 16 monomer units were replaced with a sequence derived from Y chromosome alpha satellite. This approach allowed them to knockout any interaction between CENP-B and its biochemically defined consensus element, as well as any interactions between CENP-B and elements resembling the consensus that might
20 potentially occur *in vivo*. Confirmation of abolishment of CENP-B binding to the synthetic CENP-B box null array was shown by loss of mobility shift in a gel shift assay (Figure 2).

Constructs based on the naturally occurring, unmodified D17Z1 have been used previously to generate mitotically stable SMCs in greater than 10% of drug-resistant clones

after transfection into human HT1080 cells (Rudd et al., in press; 8, 10, 18). Here, SMCs were identified in 4 of 38 colonies (Table 1), consistent with earlier data. However, when using the CENP-B box null construct in which all CENP-B boxes had been modified, only a single clone was identified to have a putative SMC out of 40 clones screened, representing a maximum *de novo* centromere formation frequency of 2.5 % (Table 1). The fact that the observed rate of *de novo* SMC formation is low but is not zero is consistent with other reports that some alpha satellite arrays that do not contain CENP-B boxes can in fact mediate apparent SMC formation at very low frequencies (25, 18), although the possibility that these represent SMCs that have acquired endogenous centromere sequences has not been rigorously excluded. Indeed, previous data have demonstrated that the likelihood of such an acquisition event is increased when the *de novo* centromere competency of the transfected DNA is lowest, as in the case of CENP-B box null constructs (8, Rudd et al. (Nov 2003), *Mol. Cell. Bio.* **23**(21):7689-7697). The data presented herein are in agreement with those recently reported by Masumoto and colleagues, who used a similar approach to abolish CENP-B boxes in a higher-order repeat derived from chromosome 21 (15). Combined, the two studies provide strong evidence that CENP-B boxes are required generally for efficient formation of *de novo* centromeres in SMC systems.

Creation of more efficient centromere constructs by increasing the density of CENP-B boxes

Several studies have now suggested a relationship between the presence of CENP-B boxes in cloned alpha satellite and the ability to form *de novo* centromeres from BAC or

YAC vectors containing the cloned arrays (8, 10, 15-19). As an extension of the data presented above and by Ohzeki et al. (15), the inventors reasoned that if the density of CENP-B boxes was indeed critical for *de novo* centromere formation, it might be possible to create synthetic alpha satellite arrays with a CENP-B box density even higher than their
5 naturally occurring counterparts. These novel synthetic arrays might form a more efficient template for centromere formation *de novo* than natural arrays.

To evaluate this hypothesis, the inventors used the strategy described above to construct a synthetic D17Z1-derived alpha satellite array supersaturated with CENP-B boxes, such that each of the 16 monomers in the HOR contained a consensus CENP-B
10 box. Notably, upon introduction into HT1080 cells by transfection, these supersaturated synthetic arrays formed SMCs *de novo* more than twice as efficiently as arrays containing the natural density of CENP-B boxes (Table 1). The frequency of SMCs within any one clone was observed to vary from 10% to 100%, similar to the ranges observed in cell lines derived from transfection with the control natural arrays (8, 17). No integration events
15 were observed cytogenetically, although Southern blot data (not shown) demonstrated the presence of BAC-specific DNA.

Initial cytogenetic estimates suggested that the SMCs (from all versions of the array) are several megabases in size; hence, suggesting recombination and rearrangement events. However, further reisolation, digestion, physical resolution and characterization revealed
20 that many of the SMCs were not rearranged and were in fact, intact, and circular plasmids that maintained the precise structure of the original vector introduced into the human cell line (*see* Figures 4-6). In addition multiple examples of SMC vectors that contain cloned

human genomic DNA fragments in addition to the synthetic alpha satellite arrays and other vector sequences shown were obtained, demonstrating that vectors containing desired genomic fragments of interest may be introduced into human cells, with the result that mitotically stable SMCs are formed that are unrearranged from the original vector
 5 sequence. In all cases, SMCs were shown to be mitotically stable in the absence of selection for a minimum of six weeks and to bind the centromere-specific protein CENP-C.

TABLE ONE

10 Effect of CENP-B box density on efficiency of SMC formation

Construct	CENP-B box density	Experiments (no.)	Clones screened (no.)	Clones with SMC (no.)	SMC formation frequency
Natural D17Z1	5/16	6	38	4	10.5 %
All CENP-B box+	16/16	15	45	10	22 %
CENP-B box null	0/16	10	40	1	2.5 %

The foregoing embodiments and advantages are merely exemplary and are not to be
 15 construed as limiting the present invention. The present teaching can be readily applied to other types of artificial chromosomes. The description of the present invention is intended

to be illustrative, and not to limit the scope of the claims. Many alternatives, modifications, and variations will be apparent to those skilled in the art.

For example, another embodiment includes the introduction of sequences that facilitate the packaging of a SMC vector into a modified viral delivery system, such as the HSV-1 amplicon systems that have been described previously (see for example Chiocca et al, and Federoff et al). These SMC vectors would include, for example, the elements described herein, such as the synthetic (CENP-B box enriched) or natural alpha satellite arrays, optionally a gene (or genes) of interest, including elements that control gene expression, and if desired, additional segments of cloned genomic DNA, which may be derived from human genomic DNA or another desired species. In the example described, in order to facilitate packaging of the vector into the viral particles used for delivery, the vector should also contain elements that facilitate such packaging, such as the HSV-1 viral packaging (pac) sequence, and replication origin (oriS) as are defined in the literature.

REFERENCES

1. Sullivan BA, Blower MD, Karpen GH. (2001) Determining centromere identity: cyclical stories and forking paths. *Nat Rev Genet.*, 2(8):584-96
2. Lee C, Wevrick R, Fisher RB, Ferguson-Smith MA, Lin CC. Human centromeric DNAs. (1997) *Hum Genet.*, 100; 291-304
3. Choo KH, Vissel B, Nagy A, Earle E, Kalitsis P. (1991) A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res.*, 19(6):1179-82
4. Willard HF. (1991) Evolution of alpha satellite. *Curr Opin Genet Dev.*, 1(4):509-14
5. Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC. (1997) The centromere: hub of chromosomal activities. *Science*, 270; 1591-1594

6. Amor DJ, Choo KH. (2002) Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet.*, 71(4):695-714
7. Saffery R and Choo KH. (2002) Strategies for engineering human chromosomes with therapeutic potential. *J. Gene Med.*, 4; 5-13
- 5 8. Grimes BR, Rhoades AA, Willard HF. (2002) Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. *Mol Ther.*, 5(6):798-805
9. Willard HF. (2001) Neocentromeres and human artificial chromosomes: an unnatural act. *Proc Natl Acad Sci U S A.*, 98(10):5374-6
- 10 10. Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF. (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet.*, 15(4):345-55
11. Waye JS, Willard HF. (1986) Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol Cell Biol.*, 6(9):3156-65
- 15 12. Willard HF. (2000) Genomics and gene therapy. Artificial chromosomes coming to life. *Science*, 290(5495):1308-9
13. Mejia JE, Willmott A, Levy E, Earnshaw WC, Larin Z. (2001) Functional complementation of a genetic deficiency with human artificial chromosomes. *Am J Hum Genet.*, 69(2):315-26
- 20 14. Grimes BR, Schindelhauer D, McGill NI, Ross A, Ebersole TA, Cooke HJ (2001) Stable gene expression from a mammalian artificial chromosome. *EMBO Rep.* 2(10):910-4
- 25 15. Ohzeki J, Nakano M, Okada T, Masumoto H. (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol.*, 159(5):765-75
16. Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF. (2001) Genomic and genetic definition of a functional human centromere. *Science.* 294(5540):109-15

17. Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat Biotechnol.*, 16(5):431-9
18. Mejia JE, Alazami A, Willmott A, Marschall P, Levy E, Earnshaw WC, Larin Z. (2002) Efficiency of de novo centromere formation in human artificial
5 chromosomes. *Genomics*, 79(3):297-304
19. Masumoto H, Ikeno M, Nakano M, Okazaki T, Grimes B, Cooke H, Suzuki N. (1998) Assay of centromere function using a human artificial chromosome. *Chromosoma*, 107(6-7):406-16
- 10 20. Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol.*, 116(3):585-96
21. Cooke CA, Bernat RL, Earnshaw WC. (1990) CENP-B: a major human centromere protein located beneath the kinetochore. *J Cell Biol* 110(5):1475-88
- 15 22. Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol.*, 109(5):1963-73
23. Earnshaw WC, Sullivan KF, Machlin PS, Cooke CA, Kaiser DA, Pollard TD, Rothfield NF, Cleveland DW. (1987) Molecular cloning of cDNA for CENP-B, the major
20 human centromere autoantigen. *J Cell Biol.*, 104(4):817-29
24. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiiri Y, Simon M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A.*, 89(18):8794-7
25. Kouprina N, Ebersole T, Koriabine M, Pak E, Rogozin IB, Katoh M, Oshimura M, Ogi K, Peredelchuk M, Solomon G, Brown W, Barrett JC, Larionov V. (2003) Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes. *Nucleic Acids Res.*, 31(3):922-34
26. Grimes BR, Warburton PE, Farr CJ. (2002) Chromosome engineering: prospects for gene therapy. *Gene Ther.*, 9(11):713-8

27. Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y. (2001) Alpha-satellite DNA of primates: old and new families. *Chromosoma*, 110(4):253-66.
28. Kipling D, Warburton PE. (1997) Centromeres, CENP-B and Tigger too. *Trends Genet.*, 13(4):141-5
- 5 29. Goldberg IG, Sawhney H, Pluta AF, Warburton PE, Earnshaw WC. (1996) Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. *Mol Cell Biol.* 16(9):5156-68
30. Kapoor M, Montes de Oca Luna R, Liu G, Lozano G, Cummings C, Mancini M, Ouspenski I, Brinkley BR, May GS. (1998) The cenpB gene is not essential in mice. *Chromosoma.*, 107(8):570-6
- 10 31. Perez-Castro AV, Shamanski FL, Meneses JJ, Lovato TL, Vogel KG, Moyzis RK, Pedersen R. (1998) Centromeric protein B null mice are viable with no apparent abnormalities. *Dev Biol.*, 201(2):135-43
32. Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, Hill J, Wreford NG, 15 de Kretser DM, Cancilla MR, Howman E, Hii L, Cutts SM, Irvine DV, Choo KH. (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *J Cell Biol.*, 141(2):309-19
33. Irelan JT, Gutkin GI, Clarke L. (2001) Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics*, 20 157(3):1191-203
34. Choo KH. (1997) Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am J Hum Genet.*, 61(6):1225-33
35. Earnshaw WC, Ratrie H 3rd, Stetten G. (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. 25 *Chromosoma*, 98(1):1-12
36. Yoda K, Ando S, Okuda A, Kikuchi A, Okazaki T. (1998) In vitro assembly of the CENP-B/alpha-satellite DNA/core histone complex: CENP-B causes nucleosome positioning. *Genes Cells*, 3(8):533-48

37. Warburton PE, Waye JS, Willard HF. (1993) Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. *Mol Cell Biol.* 13(10):6520-9.